"On the Handicapping of the First Born," A Criticism of Professor Pearson's 1914 Memoir

---

Reprinted from the Quarterly Publication of the
AMERICAN STATISTICAL ASSOCIATION,
December, 1915

BY
LOUIS I. DUBLIN, PH.D., Statistician
AND
HARRY LANGMAN, Statistical Bureau
Metropolitan Life Insurance Company, New York
1915

"ON THE HANDICAPPING OF THE FIRST BORN," A CRITICISM OF PROFESSOR PEARSON'S 1914 MEMOIR.*

In his paper "A First Study of the Statistics of Pulmonary Tuberculosis,"† 1907, Professor Pearson concluded among other things that the early born, particularly the first, are more liable to develop tuberculosis than are the later born. This finding of Pearson's was severely criticized by a number of workers, but especially by Yule and Greenwood,‡ Weinberg§ and Macaulay.|| In his more recent memoir, "On the Handicapping of the First Born," 1914, Professor Pearson elaborates this thesis and responds to the criticism directed at his first paper.

In his 1907 paper, Pearson classified the 381 patients of a tuberculosis sanatorium by order of birth. He also arranged the sibships, i. e. the group composed of the 381 tuberculous persons and their brothers and sisters, in a similar order. He then assumed that, in principle, these two distributions by order of birth should be identical, supposing that the distribution of the sibships is equivalent to the distribution by order of birth of the tuberculous population at large. He found actually that the first and second born among the patients were represented considerably above the number expected from the distribution of the sibships. He concluded, therefore, that the early born are more liable to tuberculosis.

It is this assumption of the identity of the two distributions

* Eugenics Lecture Series, X. Dulau and Company, London.

† Studies in National Deterioration, Drapers' Company Research Memoirs. Dulau and Company, London.

‡ On the Determination of the Size of Family and of the Distribution of Characters in Order of Birth from Samples taken through Members of the Sibships. Journal of the Royal Statistical Society, Vol. LXXVII, 1913–1914.

§ "Die rassenhygienische Bedeutung der Fruchtbarkeit." Arch. f. Rassen- u. Gesellschafts-Biologie, 1910, Vol. VII, p. 684. "Zur Frage der Messung der Fruchtbarkeit," *Ibid.*, 1913, Vol. X, p. 162.

|| The Supposed Inferiority of First and Second Born Members of Families—Statistical Fallacies. 1912. T. B. Macaulay, Montreal, Can.

that is the basis of much of the criticism. We shall also address ourselves to this matter below. In addition other more general criticisms have been made, especially by Macaulay, which indicate why we may expect the early born to be represented in larger proportions without involving Pearson's conclusion. They, the early born, have had a greater opportunity to pass through the so-called "danger zone" and to become patients of tuberculosis sanatoria; younger brothers and sisters may well have a pronounced tubercular diathesis, but because of their youth may not as yet have developed the disease sufficiently to need sanatorium care. There is also a tendency, even if not a very pronounced one, in tuberculous stocks for the undue representation of early borns; for the number of offspring in the family is likely to be cut down below the average through the early death of tuberculous parents, or through other circumstances which commonly break up family ties in tuberculous stock.

Pearson, in his second paper, virtually waives the above objections. He proceeds rather, at great pains, to substantiate the other assumption which was involved in his first paper, and which was criticized rather severely by Yule and Greenwood, namely, that the distribution of the sibships by order of birth is equivalent to the distribution of the tuberculous population at large. For, if this assumption is invalid, the two distributions referred to above are not comparable, and no valid conclusions can be drawn from any such comparison. In his reply, the author does not mathematically establish his conclusion as to the identity of the two distributions. He explains that he has considered the distributions of the affected in a number of cases where the distributions of the pathologic communities from which they were chosen were approximately known, and has compared the latter distributions with those obtained from the distributions of the affected by various methods, including those of his critics. On the basis of this comparison, the author has satisfied himself that the distribution of the sibships of the affected is in general the best representation of that of the pathologic community at large.

The author then extends his argument by introducing a

discussion of the mortality and "delicacy" rates of infants by order of birth. He also considers various indications of degeneracy in families by order of birth, including such traits as imbecility, epilepsy, insanity, albinism, criminality, tuberculosis, and congenital cataract. Using the same methods as have already been described for tuberculosis, he finds a marked weighting of the first born in all these defects, except in epilepsy.

After summing up his data, the author comes to the definite conclusion that "the small family is detrimental to race progress." This conclusion he makes without reservation; for such families are constituted to a greater degree by the early born who, as he maintains, are more often defective.

For the sake of the argument let us assume that the author's tables actually show what he claims for them. Then all that follows is that the early born of the community are more likely to be inferior than the later born. It does not follow, as Pearson infers, that the early born are necessarily inferior to other siblings of their own families. Nor is it true for the pathologic community at large, since the author's subjects, taken from institutions, are distributed mostly between certain ages. Thus, what may be true of patients in a tuberculosis sanatorium need by no means be true of tuberculous persons in general, who include many children, and these are rarely found in sanatoria. Especially in the face of evidence of inferiority of the last born, it is not safe to assume that superior individuals would be obtained by increasing the size of the families in question. As has already been pointed out, this inferiority of the early born may be due to an undue representation of small degenerate families; or it may be due to the higher average age of the elder born, the younger born not having as yet attained an age at which criminal or pathologic tendencies will have had opportunity for expression to a like degree; or to such circumstances as economic pressure upon older children, or to other external conditions, and not simply to the physiologic fact of being elder born.

Thus far we have assumed that Professor Pearson's method of procedure is valid. His method is based unequivocally on the assumption that the distribution according to order of

birth of the pathologic community from which his "marked" or affected subjects are obtained is identical with the distribution of the sibships of these subjects. For if that be the case he can use the distribution of the sibships of the affected as a norm in comparing with it the distribution of the affected, in the effort to show that actually the early born among his subjects preponderate beyond all expected proportions. We shall endeavor to show that, when there is no weighting according to order of birth among the individuals affected, the distribution of the affected or that of the pathologic community represented by them is not in any case comparable with that of their sibships. We propose to take the distribution of a normal population, and, supposing all members of it to be liable to some disease in equal proportions, obtain from it the distribution of the sibships of the affected by order of birth which is to be expected on the assumption made. We shall find that the distribution of the sibships is by necessity so different as to account for practically the whole difference found by Pearson. To illustrate our point, we submit the following table:

TABLE I

RECONSTRUCTION OF THE DISTRIBUTION, BY ORDER OF BIRTH, OF THE SIBSHIPS OF THE AFFECTED FROM THAT OF THE AFFECTED, OR FROM THAT OF THE GENERAL PATHOLOGIC COMMUNITY.

Sibling's Order. — Number in Family.	Industrial Classes, N. S. W.	Number of Families with 1, 2, or More in Family.	Individuals in Families of Respective Order.	Distribution of Sibships of Affected,	Per Mille Distribution of Affected.	Per Mille Distribution of Sibships of Affected.
	I	II	III	IV	V	VI
1.....	7,670	667	667	46,325	165.57	124.68
2.....	7,003	819	1,638	45,658	151.17	122.88
3.....	6,184	742	2,226	44,020	133.49	118.48
4.....	5,442	734	2,936	41,794	117.47	112.48
5.....	4,708	722	3,610	38,858	101.63	104.58
6.....	3,986	720	4,320	35,248	86.04	94.87
7.....	3,266	673	4,711	30,928	70.50	83.24
8.....	2,593	666	5,328	26,217	55.97	70.56
9.....	1,927	522	4,698	20,889	41.60	56.22
10.....	1,405	497	4,970	16,191	30.33	43.58
11.....	908	362	3,982	11,221	19.60	30.20
12.....	546	235	2,820	7,239	11.79	19.48
13.....	311	144	1,872	4,419	6.71	11.89
14.....	167	167	2,547	2,547	8.12	6.85
Over 14.....	209					
Total.....	46,325	7,670	46,325	371,554	1000	1000

In the foregoing table we give in Column I the distribution, by order of their birth, of the siblings of a community of 7,670 fertile families taken from the industrial class in New South Wales, as quoted in Pearson's paper on pulmonary tuberculosis. We suppose, also, that all of these siblings are equally liable to some disease irrespective of their order of birth. On this assumption, we shall obtain (Column VI) the distribution of the sibships of the affected which are to be expected from the distribution in Column I. From Column I we first obtain Column II, the distribution of the families in the community according to size. From Column I we have, for instance, 4,708 fifth born and 5,442 fourth born. But we must have a fourth born for every fifth born, and any excess of the former must therefore be due to families of four only, in this case 734 in Column II. In similar manner we obtain each of the other figures in Column II. Now we have, for instance, 734 families of four each. Then we have 2,936 individuals (Column III) all belonging to families of size four. Thus we obtain for Column III a distribution of the siblings of our population classified according to size of family to which they belong. Since we have assumed all these siblings equally liable to a disease, the figures of Column III must give, in their proportions, the distribution of the affected as well, according to size of family to which they belong.

We must realize now that the subjects under Professor Pearson's consideration (distributed as in Column III) are all included in the same institution. These subjects were presumably drawn at random from the pathologic community at large. We inquire now as to the probable number of families of the community we are considering that have more than one representative among the inmates of the institution. The probability of a particular member of the community being an inmate of the institution at any given time is small. The likelihood that two members of the same family are represented among our subjects is then extremely small, for, in the first place, both must be diseased; further, both must have the disease developed to the stage where they need institutional care; both must be sent to some institution; of all institutions, both must be committed to the par-

ticular one we are examining; both must be inmates of this institution at the same time, viz., that of our observation. We infer, then, that a very small number of the sibships of the affected are represented by more than one member in this particular institution. Hence we shall assume that for our purposes each of our affected comes from a distinct family, the error being negligibly small.

That being the case, Column III, besides giving the distribution (proportionately) of the affected in the institution according to size of family to which they belong, also gives the distribution of their families according to size. Let us suppose for a moment that the actual figures are magnified proportionately, and that the figures of Column III actually represent the number of the sibships of various sizes. In that case we should have, then, 46,325 families; hence 46,325 first borns. Again, there being a second born in every family of two or more, we add all the numbers of Column III from the bottom up to the second line and obtain, for the second figure of Column IV, 45,658 as the number of second borns. Similarly, we add all of Column III up to the third line and obtain 44,020 for the third figure of Column IV. In analogous manner we complete Column IV. These figures represent, then (magnified in proportion), the distribution by order of birth of the sibships of the affected. Column VI gives 1,000 individuals distributed as are the numbers of Column IV. Now, since all siblings of our community were supposed equally liable to the disease, Column I, besides giving the distribution of our community, also gives the proportionate distribution of those affected. In Column V we have 1,000 individuals distributed as the numbers of Column I. We have, then, in Column V, the per mille distribution of the diseased, as well as that of the community at large, and in Column VI the per mille distribution of the sibships of the diseased.

The discrepancy between the last two columns is striking. Thus, as to first born, the two columns give 165.6 and 124.7 as the respective numbers of affected and their sibships on a per mille basis. A like difference is found for second and third borns. These differences in proportionate distribution were obtained on the assumption that siblings of all orders of

birth were equally liable to disease—that is, that there is no weighting on the early borns. It is not surprising, then, that Professor Pearson should obtain such differences in his own tabulations, and any conclusions based on such a discrepancy fall to the ground.

We shall further illustrate our point by applying formally the method used in obtaining the above table to the distribution of the 381 patients of a sanatorium used by Pearson in his 1907 paper. We obtain the following table, where Roman numerals denote columns similar to those in Table I:

TABLE II

RECONSTRUCTION OF THE DISTRIBUTION OF THE SIBSHIPS OF 381 TUBERCULOUS INDIVIDUALS IN A SANATORIUM FROM THAT OF THE TUBERCULOUS INDIVIDUALS THEMSELVES, AND A COMPARISON WITH THE SIBSHIP DISTRIBUTION AS ACTUALLY FOUND BY PEARSON.

Sibling's Order. — Number in Family.	Distribution of Affected in Sanatorium.	Number of Families of Respective Sizes Represented by I.	Number of Affected Belonging to Families of Various Sizes.	Expected Distribution of the Sibships of the 381 Affected.	381 Distributed as the Sibships of the Affected.	Distribution of the Sibships of the Affected as Actually Found by Pearson.
	I and V	II	III	IV	VI	VII
1.....	113	34	34	381	69.69	67.1
2.....	79	38	76	347	63.47	64.4
3.....	41	−11	−33	271	49.57	58.5
4.....	52	13	52	304	55.60	50.9
5.....	39	21	105	252	46.09	43.5
6.....	18	0	0	147	26.89	32.6
7.....	18	9	63	147	26.89	22.2
8.....	9	6	48	84	15.36	15.1
9.....	3	0	0	36	6.58	10.0
10.....	3	0	0	36	6.58	6.2
11.....	3	2	22	36	6.58	3.7
12.....	1	0	0	14	2.56	2.6
13.....	1	0	0	14	2.56	1.6
14.....	1	1	14	14	2.56	1.1
Total....	381	113	381	2,083	381	381 (a)

(a) This total embraces several siblings of an order of birth beyond 14.

Here Column I or its equivalent, Column V, gives the distribution of the 381 subjects. Column VI gives the distribution of the sibships of these subjects that is to be expected from Column I. Column VII gives the distribution of the sibships of the affected as actually found by Pearson. The distributions of Columns VI and VII are fairly equivalent. But VI was obtained on the assumption that all siblings were

equally liable to the disease. Hence the discrepancy that Pearson found was to be expected on the assumption that there be no weighting according to order of birth. Certainly, then, Professor Pearson cannot use these distributions as a basis for the contention that the early born are especially liable to be diseased.

Again, let us reverse our procedure to see what number of affected may be expected from the actual distribution of the sibships. In Table I we obtained the distribution VI from that of I or V in a very definite way. Given the distribution VI, that of V is then definitely determined. Using Pearson's figures again, we obtain the following table:

TABLE III

RECONSTRUCTION OF THE DISTRIBUTION OF THE AFFECTED IN A SANATORIUM FROM THAT OF THE SIBSHIPS OF THOSE ACTUALLY AFFECTED, AND A COMPARISON WITH THE DISTRIBUTION OF THE ACTUALLY AFFECTED AS FOUND BY PEARSON.

Sibling's Order.	Distribution of the Sibships of the Affected as Found by Pearson.	Distribution of the Affected to be Expected from that of their Sibships.	Distribution of the 381 Affected as Actually Found by Pearson.
	IV and VI.	I and V	VIII
1	381	93.92	113
2	366	78.92	79
3	332	61.92	41
4	289	47.59	52
5	247	37.09	39
6	185	24.69	18
7	126	14.86	18
8	86	9.15	9
9	57	5.52	3
10	35	3.08	3
11	21	1.68	3
12	15	1.13	1
13	9	.63	1
14	6	.40	1
15 and over	9		
Total	2,164	381	381

Column IV (VI) gives the distribution of the 2,164 sibships of the affected as actually found by Pearson. Column I (V) gives the expected distribution of the affected themselves, and is obtained from VI by reversing the procedure in the other tables. Column VIII gives the actual distribution of Professor Pearson's subjects. Comparing the last two columns, we note that the first, fourth and fifth born predominate. Considered as a whole, however, the variations are too irregu-

lar to warrant any definite conclusion. The same chance variation that made 52 follow 41 in Column VIII may have raised the first figure to 113.

Professor Pearson may point out that, in some instances, he has found variations beyond even what we should anticipate on our assumptions. But we have seen that there are several reasons why we should *a priori* expect a certain predominance of first born among the marked (except, of course, in cases of congenital cataract, albinism, etc., characteristics acquired at birth). Proper allowance must be made for such factors before we can accept any decided difference in our order of birth distributions as really significant.

On reviewing Professor Pearson's data in the light of the foregoing, we find that his hypothesis is anything but substantiated, and that the cry against the pioneer child—at any rate, so far as Professor Pearson's data indicate—has little, if any, justification. It would seem obligatory that the proper method of procedure to be pursued in an investigation such as that which is proposed should be to consider only families of the same size, and that then only families with completed pathologic history should be included. This would eliminate the complication of including a number of early borns of degenerate stocks with low fertility. Again, if any definite results were obtained, we should have some reason to suppose them true for members of the same sibship. Weighting of large families as such (the primary source of the differences in our distributions) would be eliminated. Inadequacies in the records due to omission of the data concerning the ultimate history of young siblings with distinct pathologic diatheses would be accounted for. Any other like pitfalls or sources of inaccuracy must be done away with, or allowed for, before one can feel any assurance that his conclusions, if he then obtain any, include an element of truth.